

## VU Research Portal

### Mesure de l'emploi temporaire. Les données d'enquête ou de registre disent-elles la vérité ?

Pavlopoulos, Dimitris; Vermunt, J.K.

#### ***published in***

Survey Methodology  
2015

#### ***document version***

Publisher's PDF, also known as Version of record

#### ***document license***

Article 25fa Dutch Copyright Act

[Link to publication in VU Research Portal](#)

#### ***citation for published version (APA)***

Pavlopoulos, D., & Vermunt, J. K. (2015). Mesure de l'emploi temporaire. Les données d'enquête ou de registre disent-elles la vérité ? *Survey Methodology*, 41(1), 205-224. <https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/2015001/article/14151-fra.pdf?st=RALujnak>

#### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

#### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

#### **E-mail address:**

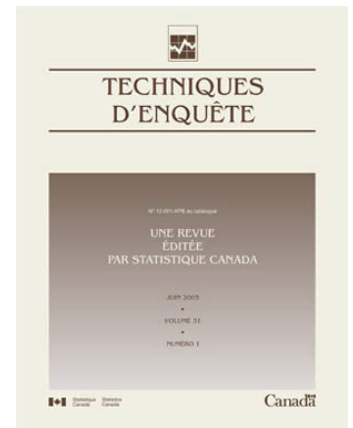
[vuresearchportal.ub@vu.nl](mailto:vuresearchportal.ub@vu.nl)

## Techniques d'enquête 41-1

# Mesure de l'emploi temporaire. Les données d'enquête ou de registre disent-elles la vérité ?

par Dimitris Pavlopoulos et Jeroen K. Vermunt

Date de diffusion : le 29 juin 2015



Statistique  
Canada

Statistics  
Canada

Canada

---

## Comment obtenir d'autres renseignements

Pour toute demande de renseignements au sujet de ce produit ou sur l'ensemble des données et des services de Statistique Canada, visiter notre site Web à [www.statcan.gc.ca](http://www.statcan.gc.ca).

Vous pouvez également communiquer avec nous par :

**Courriel** à [infostats@statcan.gc.ca](mailto:infostats@statcan.gc.ca)

**Téléphone** entre 8 h 30 et 16 h 30 du lundi au vendredi aux numéros sans frais suivants :

- |   |                |
|---|----------------|
| • Service de renseignements statistiques                                    | 1-800-263-1136 |
| • Service national d'appareils de télécommunications pour les malentendants | 1-800-363-7629 |
| • Télécopieur   | 1-877-287-4369 |

**Programme des services de dépôt**

- |                             |                |
|-----------------------------|----------------|
| • Service de renseignements | 1-800-635-7943 |
| • Télécopieur               | 1-800-565-7757 |

## Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois. À cet égard, notre organisme s'est doté de normes de service à la clientèle que les employés observent. Pour obtenir une copie de ces normes de service, veuillez communiquer avec Statistique Canada au numéro sans frais 1-800-263-1136. Les normes de service sont aussi publiées sur le site [www.statcan.gc.ca](http://www.statcan.gc.ca) sous « À propos de nous » > « Notre organisme » > « [Offrir des services aux Canadiens](#) »

## Note de reconnaissance

Le succès du système statistique du Canada repose sur un partenariat bien établi entre Statistique Canada et la population du Canada, les entreprises, les administrations et les autres organismes. Sans cette collaboration et cette bonne volonté, il serait impossible de produire des statistiques exactes et actuelles.

## Signes conventionnels dans les tableaux

Les signes conventionnels suivants sont employés dans les publications de Statistique Canada :

- . indisponible pour toute période de référence
- .. indisponible pour une période de référence précise
- ... n'ayant pas lieu de figurer
- 0 zéro absolu ou valeur arrondie à zéro
- 0<sup>s</sup> valeur arrondie à 0 (zéro) là où il y a une distinction importante entre le zéro absolu et la valeur arrondie
- <sup>p</sup> provisoire
- <sup>r</sup> révisé
- x confidentiel en vertu des dispositions de la *Loi sur la statistique*
- <sup>E</sup> à utiliser avec prudence
- F trop peu fiable pour être publié
- \* valeur significativement différente de l'estimation pour la catégorie de référence ( $p < 0,05$ )

Publication autorisée par le ministre responsable de Statistique Canada

© Ministre de l'Industrie, 2015

Tous droits réservés. L'utilisation de la présente publication est assujettie aux modalités de l'[entente de licence ouverte](#) de Statistique Canada.

Une [version HTML](#) est aussi disponible.

*This publication is also available in English.*

---

# Mesure de l'emploi temporaire. Les données d'enquête ou de registre disent-elles la vérité ?

Dimitris Pavlopoulos et Jeroen K. Vermunt<sup>1</sup>

## Résumé

L'une des principales variables de l'Enquête sur la population active des Pays-Bas est celle indiquant si un enquêté possède un emploi permanent ou temporaire. Le but de notre étude est de déterminer l'erreur de mesure de cette variable en appariant l'information tirée de la partie longitudinale de cette enquête à des données de registre uniques provenant de l'organisme de gestion des assurances sociales pour salariés des Pays-Bas (UUV). Contrairement aux approches antérieures visant à comparer des ensembles de données de ce genre, nous tenons compte du fait que les données de registre contiennent aussi des erreurs et que l'erreur de mesure qu'elles présentent est vraisemblablement corrélée dans le temps. Plus précisément, nous proposons d'estimer l'erreur de mesure dans ces deux sources en utilisant un modèle de Markov caché étendu au moyen de deux indicateurs observés du type de contrat d'emploi. Selon nos résultats, aucune des deux sources ne doit être considérée comme étant exempte d'erreur. Pour les deux indicateurs, nous constatons que les travailleurs titulaires d'un contrat d'emploi temporaire sont souvent classés incorrectement comme ayant un contrat d'emploi permanent. En particulier, dans le cas des données de registre, nous observons que les erreurs de mesure sont fortement autocorrélées, car les erreurs commises à une période ont tendance à se répéter. En revanche, lorsque l'enregistrement est correct, la probabilité qu'une erreur soit commise à la période suivante est presque nulle. Enfin, nous constatons que les contrats d'emploi temporaire sont plus répandus que ne le laisse supposer l'Enquête sur la population active, tandis que les taux de transition entre les contrats d'emploi temporaire et permanent sont nettement moins élevés que ne le suggèrent les deux ensembles de données.

**Mots-clés :** Contrats d'emploi temporaire; erreur de mesure; modèle de Markov caché; données de registre.

## 1 Introduction

La question de l'emploi temporaire occupe une place de plus en plus importante dans le débat économique et politique. Les contrats d'emploi temporaire permettent aux employeurs de contourner les règlements stricts en matière d'embauche et de licenciement (Bentolila et Bertola 1990; Booth 1997; Cahuc et Postel-Vinay 2002) et parfois même les règlements concernant la rigidité des salaires (OECD 2002). Surtout durant les récessions économiques, les employeurs recourent aux contrats d'emploi temporaire pour ajuster leur effectif en fonction des fluctuations de la demande de produits.

Les Pays-Bas font figure de pionniers en matière de souplesse d'emploi depuis le début des années 1990. La souplesse contractuelle est une caractéristique importante du marché du travail néerlandais. L'emploi temporaire a augmenté fortement, pour passer de 5,9 % en 1991 à 17,1 % en 2010 (OECD 2012), tandis que la contribution de la croissance de l'emploi temporaire à la croissance de l'emploi total a été de 9,9 points de pourcentage de 1990 à 2000 (OECD 2002). Les employeurs adoptent habituellement une stratégie d'effectif à « capacité minimale » (Sels et Van Hootegeem 2001), qui consiste à offrir des contrats permanents à leurs travailleurs « de base » et des contrats temporaires aux autres travailleurs afin de pouvoir ajuster leur effectif en période de ralentissement économique.

Jusque récemment, les statistiques sur les contrats d'emploi temporaire aux Pays-Bas reposaient exclusivement sur des données d'enquêtes auprès des ménages et d'enquêtes sur la population active, mais

1. Dimitris Pavlopoulos, Université libre d'Amsterdam, département de sociologie, De Boelelaan 1081, 1081 HV Amsterdam, Pays-Bas. Courriel : d.pavlopoulos@vu.nl; Jeroen K. Vermunt, Université de Tilburg, département de méthodologie et de statistique, CP 90153, 5000 LE Tilburg, Pays-Bas. Courriel : j.k.vermunt@tilburguniversity.edu.

il existe aujourd'hui des données de registre de haute qualité qui peuvent compléter les données d'enquête, voire les remplacer. La première comparaison des deux sources de données a révélé certains chiffres gravement divergents quant à la taille de l'emploi temporaire. En 2009, la part de tous les types de contrats d'emploi temporaire était de 15,4 % selon l'Enquête sur la population active (EPA), tandis qu'elle était de 23,6 % selon les données du « Polisadministratie » (PA), qui sont des données de registre fournies par l'organisme de gestion des assurances sociales pour salariés (UWV) (Hilbers, Houwing et Kösters 2011). Comme la taille de l'emploi temporaire est un élément très important à prendre en considération dans l'élaboration des politiques relatives au marché du travail, *Statistics Netherlands* a entrepris de résoudre les divergences entre les deux sources de données. Une analyse plus approfondie des données ne s'est pas avérée très prometteuse. Les résultats préliminaires indiquent que 15,6 % des personnes titulaires d'un contrat d'emploi permanent selon l'EPA semblent posséder un contrat d'emploi temporaire selon le PA, tandis que 18,3 % des personnes titulaires d'un contrat d'emploi temporaire d'une durée inférieure à un an selon l'EPA semblent posséder un contrat d'emploi permanent selon le PA (Mars 2011). Bien qu'une part des incohérences puisse être expliquée par les définitions un peu différentes de l'emploi temporaire dans les deux sources de données, de grands écarts persistent même quand on utilise un échantillon apparié et que l'on sélectionne les cas pour lesquels il n'existe pas de différence de définition.

Comme le laissent entendre des études antérieures, l'erreur de mesure peut expliquer les incohérences constatées entre les données d'enquête et les données de registre. Dans le cas des données d'enquête, l'erreur de mesure est reconnue comme étant une source importante de biais (Rodgers, Brown et Duncan 1993; Pischke 1995; Bollinger 1996; Rendtel, Langeheine et Berntsen 1998; Bound, Brown et Mathiowetz 2001; Biemer 2011). L'erreur de mesure relative au type de contrat d'emploi n'a fait l'objet d'aucune étude jusqu'à présent, mais des travaux de recherche sur d'autres caractéristiques du marché du travail, dont l'activité, les salaires, le nombre d'heures de travail, l'industrie et la profession, montrent que les données d'enquête peuvent contenir de grandes quantités d'erreurs de mesure, susceptibles de biaiser sévèrement les résultats des analyses statistiques. Ainsi, Biemer (2004) soutient que dans les éditions de 1992 à 1994 de la *Current Population Survey*, 20,9 % des enquêtés en chômage ont été classés incorrectement dans d'autres catégories. Gottschalk (2005) indique que les deux tiers des réductions de salaire nominal sans changement d'emploi étaient dues à des erreurs de mesure. En particulier, 17 % des travailleurs déclarent une réduction de salaire nominal d'une année à l'autre tout en restant au service du même employeur. Cependant, si l'on neutralise l'effet de l'erreur de mesure, la proportion de travailleurs qui demeurent avec le même employeur et font face à des réductions de salaire nominal annuelles ne dépasse pas 4 % à 5 %. En s'appuyant sur l'étude de validation de la *Panel Study of Income Dynamics* (PSID), Mathiowetz (1992) constate que la concordance entre les registres des entreprises et les réponses aux enquêtes pour ce qui est de la catégorie professionnelle est de 87,3 %. Brown et Medoff (1996) observent une corrélation de 0,82 entre les registres des entreprises et les réponses aux enquêtes en ce qui a trait à la taille de l'établissement, et une corrélation de 0,86 en ce qui concerne la taille de l'entreprise.

La recherche sur l'erreur de mesure est nettement moins riche dans le cas des données de registre que dans celui des données d'enquête. Les données de registre sont habituellement traitées comme étant exemptes d'erreur et utilisées comme « norme de référence » quand elles sont comparées aux données d'enquête. Ainsi, la plupart de la recherche fondée sur l'étude de validation de la PSID fait appel à cette hypothèse (Duncan et Hill 1985; Rodgers et coll. 1993; Bound, Brown, Duncan et Rodgers 1994;

Pischke 1995). Toutefois, certains travaux montrent aussi que l'hypothèse de la « norme de référence » n'est pas toujours plausible. Kapteyn et Ypma (2007) étudient l'erreur de mesure dans la rémunération et, quoiqu'ils retiennent l'hypothèse que les données de registre sont exemptes d'erreur, ils tiennent compte de l'erreur dans l'appariement des données d'enquête aux données de registre. En particulier, ils supposent qu'un enregistrement dans le registre est identique à un enregistrement dans l'enquête avec une certaine probabilité. Ils concluent que l'introduction de cette source supplémentaire d'erreur modifie le profil de l'erreur de mesure dans l'enquête. Abowd et Stinson (2005) comparent la rémunération déclarée dans le cadre de la *Survey of Income and Program Participation* (SIPP) et les données des *Detailed Earnings Records* (DER). Ils constatent que l'erreur de mesure est plus importante dans les données administratives des DER (20 % à 27 %) que dans les données de la SIPP (13 % à 15 %). En comparant les mêmes ensembles de données, Gottschalk et Huynh (2010) soutiennent que l'erreur de mesure peut biaiser fortement les mesures de l'inégalité des revenus.

L'objectif du présent article consiste à estimer la grandeur de l'erreur de mesure concernant le type de contrat d'emploi dans l'EPA des Pays-Bas. À cette fin, nous comparons les données de l'enquête à des données provenant du registre PA. Les données de registre sont traitées comme n'étant pas exemptes d'erreur, car nous modélisons simultanément l'erreur de mesure dans les deux sources. Nous utilisons un modèle de Markov caché étendu au moyen de deux indicateurs du type de contrat (temporaire ou permanent), chacun provenant de l'une de nos sources de données.

Le restant de l'article est présenté comme il suit. À la section 2, nous nous étendons davantage sur le problème de la mesure de l'emploi temporaire au Pays-Bas en présentant les renseignements pertinents sur les deux sources de données, ainsi que certaines statistiques descriptives. À la section 3, nous décrivons le modèle de Markov caché utilisé pour l'étude. À la section 4, nous discutons des résultats de l'analyse. Enfin, à la section 5, nous présentons les conclusions de notre étude.

## 2 Description des deux sources de données

Les deux sources de données sur les contrats d'emploi temporaire sont l'Enquête sur la population active (en néerlandais : *Enquête Beroepsbevolking*) administrée par *Statistics Netherlands* (en néerlandais : *Centraal Bureau voor de Statistiek* – CBS) et l'ensemble de données « Polisadministratie » de l'organisme de gestion des assurances sociales pour salariés (UWV). L'EPA est une enquête trimestrielle à panel rotatif sur les caractéristiques du marché du travail qui est représentative de la population néerlandaise de 15 ans et plus. L'enquête a été lancée en 1987, mais sa composante longitudinale a été introduite en 1999. Depuis 1999, les enquêtés sont interviewés durant cinq vagues consécutives du panel, ce qui permet d'étudier les évolutions individuelles à court terme sur le marché du travail. L'information recueillie a trait à la situation au moment de l'interview. Les interviews sont réparties assez uniformément au cours du trimestre.

Les erreurs de mesure du type de contrat dans le cadre de l'EPA sont, comme cela est habituellement le cas dans les enquêtes, le résultat d'erreurs de déclaration par les enquêtés ou d'erreurs d'enregistrement des réponses par les enquêteurs. Le recours à des interviews par personne interposée est une source supplémentaire d'erreur. Habituellement, dans l'EPA, un seul membre du ménage fournit les réponses pour tous les membres du ménage inclus dans l'échantillon, ce qui accroît l'erreur de mesure. Dans notre

échantillon de l'EPA, 40,1 % des observations proviennent d'interviews par personne interposée. Une autre cause possible d'erreur de mesure tient au fait que les travailleurs peuvent confondre le contrat d'emploi juridique et le contrat d'emploi implicite ou psychologique avec leur employeur. En particulier chez les cohortes de jeunes travailleurs parmi lesquels les contrats flexibles sont très répandus et dans les secteurs où la mobilité d'emploi est grande et où les conditions d'emploi évoluent, comme le secteur de la santé, les travailleurs peuvent déclarer qu'ils ont un contrat permanent en se basant sur les promesses de l'employeur, alors qu'en réalité, ils sont employés aux termes d'un contrat temporaire.

Le PA est un ensemble de données de registre unique contenant l'information sur le marché du travail et le revenu de tous les travailleurs assurés au Pays-Bas. Cet ensemble de données est construit en recueillant et en appariant l'information provenant de diverses sources, comme le Bureau de l'impôt (en néerlandais : *Belastingdienst*) – y compris des données provenant des déclarations de revenu des particuliers aux fins de l'impôt (en néerlandais : *jaaropgave*), les déclarations fournies par les agences d'emploi temporaire (en néerlandais : *weekaanleveringen*) et le registre de la population (en néerlandais : *Gemeentelijke BasisAdministratie persoonsgegevens* - GBA). Le PA est administré par l'organisme de gestion des assurances sociales pour salariés des Pays-Bas (UWV).

L'UWV a tout intérêt à maintenir le haut niveau de qualité et d'exactitude du PA, car cette source de données est utilisée par plusieurs institutions gouvernementales. Par exemple, les cotisations de sécurité sociale, les allocations de logement (en néerlandais : *huurtoeslag*), et les indemnités de soins de santé (en néerlandais : *zorgtoeslag*) sont déterminées en se servant de l'information provenant de cet ensemble de données. Afin d'améliorer la qualité des données, le PA a fait l'objet de plusieurs révisions depuis 2006. Il n'y a pas de données manquantes car les employeurs sont obligés de transmettre les relevés d'information fiscale. Cependant, alors que l'ensemble de données contient des renseignements mensuels, les employeurs transmettent habituellement l'information pertinente une fois par an (il est impossible d'extraire le moment de la transmission). Cette situation peut donner lieu à des erreurs pour la période comprise entre deux transmissions consécutives, surtout en ce qui concerne la mesure du type de contrat, qui n'est manifestement pas la variable la plus importante pour les utilisateurs du PA. Par conséquent, il est probable que, si une erreur est commise concernant le type de contrat, elle persistera jusqu'au moment où l'employeur soumettra le rapport suivant à l'UWV. Cela signifie que l'on peut s'attendre à ce que l'erreur de mesure dans le PA soit autocorrélée.

Pour les besoins de notre étude, nous choisissons des participants à l'EPA qui ont été interviewés pour la première fois durant le premier trimestre de 2007. Puisque nous nous concentrons sur les personnes ayant un emploi, nous avons gardé dans l'échantillon les personnes âgées de 25 à 55 ans. Après l'application de la contrainte d'âge, nous avons obtenu une taille d'échantillon de 11 632 personnes. Pour toutes ces personnes, *Statistics Netherlands* a apparié l'information provenant de l'EPA à l'information mensuelle provenant du PA en utilisant le numéro de sécurité sociale des individus. Le niveau d'appariement atteint était de 98 % et toutes les incohérences pertinentes ont été résolues (l'appariement et le contrôle de la qualité ont été effectués par *Statistics Netherlands*). Notre ensemble de données finales se présente sous la forme d'un fichier personne-mois pour 11 632 personnes avec 15 observations pour la période de janvier 2007 à mars 2008 contenant l'information complète provenant du PA et l'information partiellement observée (5 observations – une réponse tous les trois mois) provenant de l'EPA. L'ensemble de données appariées est illustré au tableau 2.1. Cet ensemble de données de panel n'est pas équilibré pour l'EPA, car nos données d'enquête souffrent d'une certaine attrition. Plus précisément, des

11 632 personnes qui avaient répondu à la première interview, 9 970 restaient dans l'échantillon de l'EPA pour la deuxième interview, 9 113 pour la troisième, 8 953 pour la quatrième et 8 629 pour la dernière interview. Dans le cas des données du PA pour cet échantillon, il n'y a pas d'attrition, de sorte que l'échantillon est entièrement équilibré.

**Tableau 2.1**  
**Une illustration de notre exemple**

<b>EPA</b>												
<b>Polisadministratie</b>												
	Janv.07	Fév. 07	Mars 07	Avr. 07	Mai 07	Jun 07	Juill. 07	Août 07	Sept. 07	Oct. 07	Nov. 07	Déc. 07
<b>EPA</b>												
<b>Polisadministratie</b>												
	Janv. 08	Fév. 08	Mars 08									

Nota : Le tableau illustre la correspondance entre le panel rotatif de l'EPA et les observations mensuelles provenant du registre Polisadministratie. Il a trait aux personnes qui ont été interviewées le premier mois de chaque trimestre. Une cellule ombrée en gris indique une observation valide.

La principale variable d'intérêt de notre étude est le type de contrat d'emploi, qui peut prendre trois valeurs : permanent, temporaire et « autre ».

Le type de contrat est déterminé d'après l'emploi principal, donc l'information sur les autres emplois qu'une personne pourrait occuper n'est pas prise en compte. Les personnes qui n'ont pas d'emploi rémunéré sont classées dans la catégorie « autre ». Il convient de souligner que cette dernière catégorie est assez hétérogène, car elle englobe, entre autre, les catégories de travail autonome, de chômage et d'études à temps plein. Cependant, il est nécessaire d'ajouter cette catégorie dans notre analyse, car dans les modèles de Markov, les classes latentes doivent être mutuellement exclusives et exhaustives.

Le tableau 2.2 présente la distribution des types de contrats observée pour le premier mois de la période de référence selon les données d'enquête et les données de registre. Les écarts les plus importants ont lieu pour les pourcentages de personnes possédant des contrats d'emploi permanent et temporaire, et sont moindre pour la catégorie « autre ». Selon les données d'enquête, en janvier 2007, 8 % de la population active détenait un contrat d'emploi temporaire, tandis que selon les données du registre, la proportion était plus élevée (12,3 %).

**Tableau 2.2**  
**Distribution des types de contrats selon l'enquête et selon le registre**

	<b>Enquête</b>	<b>Registre</b>
Permanent	0,659	0,602
Temporaire	0,080	0,123
Autre	0,261	0,275
Total	1,0	1,0
Cas	3 887	11 632

Nota : Ces distributions de fréquence ont trait au premier mois de la période de référence, janvier 2007. L'échantillon de l'EPA est plus petit que l'échantillon du PA car seulement 3 887 participants à l'EPA ont été interviewés pour la première fois en janvier 2007. Les autres participants ont été interviewés en février et en mars 2007.



Le tableau 2.3 donne le croisement du type de contrat selon les deux sources pour l'échantillon groupé. Ce tableau confirme les écarts importants entre les deux sources de données signalés par *Statistics Netherlands*. Ces écarts ont trait principalement aux personnes qui sont enregistrées comme travaillant aux termes d'un contrat temporaire. Plus précisément, 50,2 % de personnes qui, selon les données du registre, ont un contrat d'emploi temporaire semblent posséder un contrat d'emploi permanent selon l'enquête. Des divergences plus faibles, mais néanmoins existantes, se dégagent pour les personnes qui sont inscrites comme ayant un contrat d'emploi permanent ou comme appartenant à une autre catégorie.

Les incohérences de classification des personnes présentées au tableau 2.3 ont des conséquences importantes en ce qui concerne les transitions entre les différents états d'emploi. Le tableau 2.4 donne les taux de transition sur trois mois pour les cas assortis d'une observation valide provenant de l'EPA. Ce tableau indique que les données du registre contiennent un plus grand nombre de transitions que les données d'enquête. En particulier, parmi les personnes ayant un contrat d'emploi temporaire au mois  $t - 3$ , 5,7 % ont un contrat d'emploi permanent au mois  $t$  selon les données d'enquête et 8,5 % sont dans cette situation selon les données du registre.

**Tableau 2.3**  
**Tableau croisé du type de contrat selon l'enquête et selon le registre**

Données du registre	Données d'enquête			
	Permanent	Temporaire	Autre	Total
Permanent	0,944	0,039	0,017	1,0
Temporaire	0,502	0,437	0,061	1,0
Autre	0,081	0,030	0,889	1,0
Total	0,667	0,087	0,246	1,0
Cas	32 225	4 216	11 856	48 297

Nota : Les distributions de fréquence sont calculées pour l'échantillon groupé. Le total général représente le nombre d'enregistrements de l'EPA inclus dans notre analyse dans l'échantillon groupé.

**Tableau 2.4**  
**Transitions sur trois mois observées dans l'EPA et dans le PA**

Transitions observées d'après les données de l'enquête				
Contrat en $t-3$	Contrat en $t$			
		Permanent	Temporaire	Autre
	Permanent	0,981	0,009	0,010
	Temporaire	0,057	0,889	0,054
	Autre	0,017	0,035	0,948
Total		0,674	0,089	0,237
Transitions observées d'après les données du registre				
Contrat en $t-3$	Contrat en $t$			
		Permanent	Temporaire	Autre
	Permanent	0,967	0,018	0,015
	Temporaire	0,085	0,860	0,055
	Autre	0,018	0,036	0,946
Total		0,624	0,128	0,247

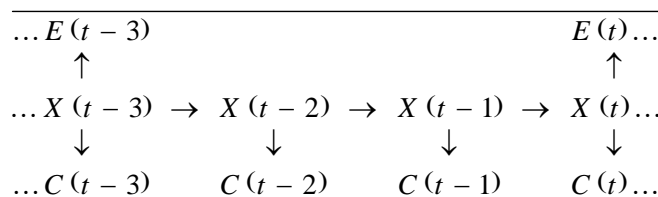
Nota : Il s'agit des taux de transition sur une période de trois mois et pour 34 820 cas de notre échantillon groupé. Ces cas proviennent des participants à l'EPA qui figurent au moins deux fois dans notre échantillon.

### 3 Modèle de Markov caché utilisé pour estimer l'erreur de mesure du type de contrat

Le modèle que nous utilisons pour estimer l'erreur dans la mesure du type de contrat est un modèle de Markov caché ou latent. Ce modèle a été utilisé pour estimer l'erreur de mesure de variables provenant des enquêtes sur l'emploi (voir, entre autre, van der Pol et Langeheine 1990; Rendtel et coll. 1998; Bassi, Hageaars, Croon et Vermunt 2000; Biemer et Bushery 2000; Biemer 2011; Pavlopoulos, Muffels et Vermunt 2012). Notre application diffère des applications susmentionnées en ce sens que nous avons deux mesures au lieu d'une seule pour la variable de résultat; autrement dit, le type de contrat d'après le registre PA et d'après l'EPA. D'autres exemples d'applications de modèles de Markov latents spécifiées pour de multiples variables de réponse figurent dans Langeheine (1994), Paas, Vermunt et Bijmolt (2007), Bartolucci, Lupporelli et Montanari (2009) et Manzoni, Vermunt, Luijkx et Muffels (2010).

Soit  $C_{it}$  et  $E_{it}$  l'état observé de la personne  $i$  au point dans le temps  $t$  selon le registre et selon l'enquête, respectivement, où  $i = 1, \dots, N$  et  $t = 0, \dots, T$ . Pour tenir compte du fait que  $E_{it}$  n'est observé que tous les trois mois, nous utilisons la variable indicatrice  $\delta_{it}$  qui est égale à 1 si l'information de l'enquête était disponible pour le mois concerné et 0 autrement. En plus des mesures provenant du registre et de l'enquête, le modèle de Markov caché contient une variable non observable représentant le type de contrat réel d'une personne au point  $t$  dans le temps. Nous désignons cet état latent par  $X_{it}$ . Notons que  $C_{it}$ ,  $E_{it}$  et  $X_{it}$  peuvent prendre trois valeurs représentant les catégories de contrat permanent, temporaire et autre. Nous désignons une catégorie particulière de ces variables par  $c_t$ ,  $e_t$ , et  $x_t$ , respectivement.

Le diagramme de cheminement pour le modèle de Markov caché d'intérêt est illustré à la figure 3.1. Pour simplifier, ce diagramme de cheminement se rapporte uniquement aux personnes qui sont entrées dans l'échantillon de l'EPA durant un mois particulier. Pour cette raison, des quatre observations qui sont illustrées dans le diagramme, seules celles des mois  $t-3$  et  $t$  ne manquent pas pour l'EPA. Comme le montre la figure, le type de contrat latent  $X_{it}$  suit un processus de Markov d'ordre 1; autrement dit, le contrat réel au point dans le temps  $t$ ,  $X_{it}$ , est indépendant du contrat au point dans le temps  $t'$ ,  $X_{it'}$ , pour  $t' < t-1$ , conditionnellement à l'état au temps  $t-1$ ,  $X_{i(t-1)}$ . Une autre hypothèse est que les états observés sont indépendants l'un de l'autre à l'intérieur et entre les points dans le temps, ce que nous appelons l'hypothèse d'indépendance locale ou l'hypothèse d'erreurs de classification indépendantes (ECI). On peut aussi constater que  $E_{it}$  est observé seulement tous les trois points dans le temps.



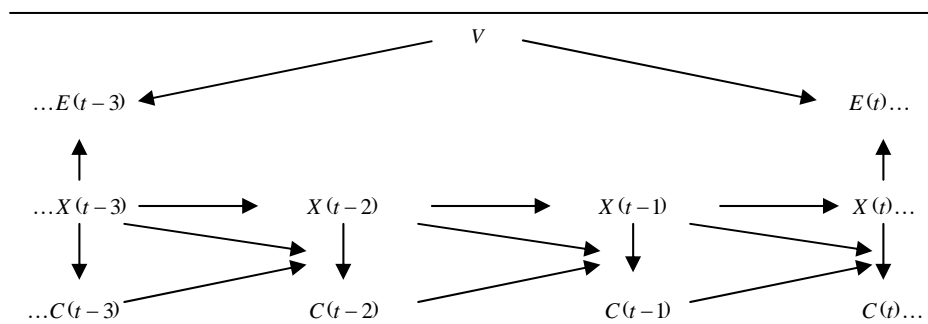
**Figure 3.1** Diagramme de cheminement pour le modèle de Markov caché avec deux indicateurs (partiellement) observés

Comme il est mentionné à la section précédente, nous utilisons les données pour 15 mois, ce qui signifie que  $t$  va de 0 à  $T = 14$ . La probabilité de suivre un certain cheminement observé sur la période de  $T + 1$  mois peut être exprimée comme il suit :

$$P(\mathbf{C}_i = \mathbf{c}_i, \mathbf{E}_i = \mathbf{e}_i) = \sum_{x_0=1}^3 \sum_{x_1=1}^3 \dots \sum_{x_T=1}^3 P(X_{i0} = x_0) \prod_{t=1}^T P(X_{it} = x_t | X_{i(t-1)} = x_{t-1}) \prod_{t=0}^T P(C_{it} = c_t | X_{it} = x_t) \prod_{t=0}^T P(E_{it} = e_t | X_{it} = x_t)^{\delta_{it}}. \quad (3.1)$$

Les probabilités pertinentes qui figurent dans cette équation sont les probabilités de l'état initial  $P(X_{i0} = x_0)$ , les probabilités de transition propre à la période  $P(X_{it} = x_t | X_{i(t-1)} = x_{t-1})$ , les probabilités d'erreur de mesure pour le registre  $P(C_{it} = c_t | X_{it} = x_t)$ , et les probabilités d'erreur de mesure pour l'enquête  $P(E_{it} = e_t | X_{it} = x_t)$ .

Jusqu'à présent, nous avons supposé que l'erreur de mesure n'était pas corrélée entre les divers points dans le temps – c'est-à-dire que l'hypothèse ECI est vérifiée – ce qui pourrait ne pas être raisonnable dans le cas de notre application. Avant tout, comme il est mentionné à la section précédente, l'erreur de mesure dans les données du registre est vraisemblablement autocorrélée; autrement dit, l'existence d'une erreur de classification entre  $X_{it}$  et  $C_{it}$  au point dans le temps  $t$  augmente la probabilité que la même erreur soit présente au point dans le temps  $t + 1$ . Cela tient au fait que les erreurs que les employeurs commettent dans leurs registres ne sont pas corrigées jusqu'à ce qu'ait lieu un contrôle ordinaire. Dans les données d'enquête, surtout parce qu'elles sont prospectives plutôt que rétrospectives, il n'y a aucune raison de penser qu'il existe une structure d'erreur autocorrélée « directe » similaire. Par contre, les erreurs dans les données d'enquête peuvent être corrélées dans le temps parce que la probabilité de commettre une erreur peut varier d'un groupe de personnes à l'autre, ce qui est parfois appelé erreur de mesure différentielle. En particulier, l'erreur de mesure dans les données d'enquête est vraisemblablement plus importante dans les secteurs où la mobilité est fréquente et où il existe une ambiguïté quant aux accords conclus entre les employeurs et les travailleurs, par exemple dans le secteur de la santé. De surcroît, les erreurs peuvent être plus importantes pour les jeunes travailleurs qui se soucient moins d'une relation de long terme avec l'employeur et qui, par conséquent, pourraient avoir une vision moins claire que les enquêtés plus âgés des modalités officielles de leur contrat. La figure 3.2 représente le diagramme de cheminement du modèle lorsque l'on corrige l'hétérogénéité et l'autocorrélation possibles de l'erreur de mesure, où  $V$  représente les variables observées qui introduisent une corrélation dans le temps de l'erreur de mesure dans les données d'enquête.



**Figure 3.2** Diagramme de cheminement pour le modèle de Markov caché avec deux indicateurs et des erreurs corrélées

Comme il est également important de tenir compte de l'hétérogénéité dans la partie structurelle d'un modèle de Markov (Shorrock 1976), le modèle est étendu au moyen de variables observées – éventuellement variant en fonction du temps – qui ont une incidence sur l'état initial et les probabilités de transition latentes, en suivant l'approche de Vermunt, Langeheine et Böckenholt (1999). Nous désignons ces variables de contrôle par  $\mathbf{Z}_{it}$ . Cependant, ces variables de contrôle observées ne peuvent pas traduire complètement l'hétérogénéité présente dans les probabilités de transition latentes, car celles-ci peuvent également être affectées par des traits personnels non observés, tels que la motivation et les compétences. En suivant l'approche la plus classique dans le cadre des modèles de Markov caché, nous corrigeons l'hétérogénéité non observée en supposant que la population est constituée d'un petit nombre de classes latentes dont l'état initial et les probabilités de transition diffèrent (Poulsen 1990). De cette façon, nous évitons les hypothèses de distribution de la variable latente peu attrayantes qui sont adoptées dans les modèles à effets aléatoires continus (Heckman et Singer 1984; Vermunt 1997). Le nombre de classes latentes  $K$  peut être déterminé en utilisant des indices d'ajustement du modèle.

Dans notre modèle de Markov caché mixte, la probabilité conjointe d'avoir un cheminement vers un état observé particulier conditionnellement aux valeurs du prédicteur peut s'exprimer par l'équation :

$$\begin{aligned}
 P(\mathbf{C}_i = \mathbf{c}_i, \mathbf{E}_i = \mathbf{e}_i | \mathbf{V}_i, \mathbf{Z}_i) &= \sum_{k=1}^K \sum_{x_0=1}^3 \sum_{x_1=1}^3 \dots \sum_{x_T=1}^3 \pi_k P(X_{i0} = x_0 | \mathbf{Z}_{i0}, k) \\
 &\quad \prod_{t=1}^T P(X_{it} = x_t | X_{i(t-1)} = x_{t-1}, \mathbf{Z}_{it}, k) \\
 &\quad P(C_{i0} = c_0 | X_{i0} = x_0) \\
 &\quad \prod_{t=1}^T P(C_{it} = c_t | X_{it} = x_t, X_{i(t-1)} = x_{t-1}, C_{i(t-1)} = c_{t-1}) \\
 &\quad \prod_{t=0}^T P(E_{it} = e_t | X_{it} = x_t, \mathbf{V}_{it})^{\delta_{it}},
 \end{aligned} \tag{3.2}$$

qui spécifie un modèle de mélange fini comportant  $K$  classes latentes pour tenir compte de l'hétérogénéité non observée dans l'état latent initial et dans les probabilités de transition latentes.  $\pi_k$  est la probabilité d'appartenir à la classe latente  $k$ ,  $\mathbf{V}_{it}$  est le vecteur de covariables ayant une incidence sur l'erreur de mesure dans les données d'enquête (âge et interview par personne interposée), et  $\mathbf{Z}_{it}$  est le vecteur des covariables ayant une incidence sur les probabilités de transition latentes (sexe, âge, niveau de scolarité et pays d'origine).  $\mathbf{Z}_{i0}$  est le vecteur des valeurs de ces covariables au point dans le temps initial.

Contrairement à l'équation 3.1, dans l'équation 3.2, les probabilités d'erreur dans les données d'enquête peuvent dépendre des covariables ( $\mathbf{V}_{it}$ ). Les effets des covariables sur ces probabilités d'erreur sont modélisés en utilisant un modèle logit. En outre, les probabilités d'erreur dans les données du registre peuvent dépendre du type de contrat observé décalé et du type de contrat réel décalé. Notons que  $X_{i(t-1)}$  et  $C_{i(t-1)}$  peuvent prendre trois valeurs, ce qui implique qu'il existe neuf (3 fois 3) ensembles différents de

probabilités d'erreur dans les données du registre, un pour chaque combinaison possible de contrat observé et de contrat latent décalés. Comme cela n'a pas de sens d'estimer toutes ces probabilités d'erreur librement, nous avons utilisé un modèle restreint. Plus spécifiquement, nous définissons un modèle logit pour  $P(C_{it} = c_t | X_{it} = x_t, X_{i(t-1)} = x_{t-1}, C_{i(t-1)} = c_{t-1})$  de la forme  $\alpha_{c_t, x_t} + \beta_{c_t, c_{t-1}, x_t, x_{t-1}}$ , où  $\beta_{c_t, c_{t-1}, x_t, x_{t-1}}$  est un paramètre libre quand  $c_t = c_{t-1} \neq x_t = x_{t-1}$  (quand la même erreur est commise entre des points dans le temps adjacents) et égal à 0 autrement. Ce modèle, qui contient six paramètres supplémentaires comparativement à un modèle sans effets décalés sur les probabilités d'erreur de classification, exprime que la vraisemblance de commettre une erreur particulière dépend du fait que *la même erreur* a été ou non commise au point dans le temps précédent. Des structures d'erreur corrélées contraintes similaires ont été utilisées par Manzoni et coll. (2010) dans un modèle de Markov latent pour des réponses recueillies rétrospectivement.

Pour restreindre l'état initial et les probabilités de transition latentes, nous nous servons aussi de modèles logit, tandis que pour les transitions latentes, nous utilisons des modèles qui séparent les coefficients selon l'état d'origine. Le même ensemble de covariables ( $\mathbf{Z}_{i0}$  et  $\mathbf{Z}_{it}$ , respectivement) est introduit dans les modèles pour l'estimation de l'état initial et des probabilités de transition latentes. Notons que le modèle de Markov caché mixte décrit dans l'équation 3.2 repose sur l'hypothèse d'un processus de Markov d'ordre 1 pour les états réels conditionnellement aux valeurs individuelles des covariables et des effets non observés constants dans le temps, mais que cette hypothèse ne doit pas être vérifiée après la marginalisation sur les valeurs des covariables et les classes latentes. Un simple modèle de Markov de premier ordre ne conviendrait pas pour les transitions dans l'emploi, particulièrement au niveau mensuel, parce qu'il existe une dépendance à l'égard de la durée dans le chômage. Par exemple, il est peu probable qu'une personne en chômage durant les mois 3 à 9 ait la même probabilité de se trouver dans un état particulier sur le marché du travail au mois 10 qu'une personne qui n'a été en chômage que durant le mois 9. Cependant, dans un modèle de Markov caché, le biais de l'erreur de classification due à la violation de l'hypothèse de Markov est minime. En utilisant des simulations, Biemer et Bushery (2000) montrent que, même dans les cas d'une violation grave de l'hypothèse de Markov, dans un modèle de Markov caché, le biais de l'estimation de l'erreur de classification dans la catégorie du chômage ne dépasse pas 3 %.

Les estimations du maximum de vraisemblance des paramètres du modèle sont obtenues en utilisant une variante de l'algorithme espérance-maximisation (EM) (Dempster, Laird et Rubin 1977) appelée algorithme *forward-backward* ou algorithme de Baum-Welch (Baum, Petrie, Soules et Weiss 1970). Nous utilisons une extension de cet algorithme pour modèles de Markov latents mixtes avec covariables comme il est décrit, entre autres, dans Vermunt, Tran et Magidson (2008) et Pavlopoulos et coll. (2012). À l'étape E, nous calculons la log-vraisemblance prévue pour les données complètes, ce qui comprend le calcul des probabilités a posteriori marginales pertinentes pour les classes latentes et les états latents. À l'étape M, les paramètres du modèle sont mis à jour en utilisant des algorithmes standard pour l'analyse par régression logistique, où les probabilités a posteriori marginales sont utilisées comme pondérations. Cet algorithme est implémenté dans le programme Latent GOLD (Vermunt et Magidson 2008), qui fournit aussi les erreurs-types pour les paramètres du modèle (d'autres programmes populaires pour l'estimation des modèles de Markov latents sont MPLUS, LEM et PANMARK).

Les valeurs manquantes en raison de la conception de l'enquête (parce que les participants ne sont interviewés qu'une fois tous les trois mois) manquent complètement au hasard (MCAR pour *Missing*

*Completely At Random*). Les valeurs manquantes dues à l'attrition sont traitées comme des données manquant au hasard (MAR pour *Missing At Random*). Plus précisément, de la manière standard dans la procédure d'estimation du MV, nous maximisons la log-vraisemblance pour les données observées incomplètement, ce qui s'obtient en éliminant par intégration les valeurs manquantes. Cette procédure est valide sous MAR.

Comme le plan de sondage de l'EPA est complexe, nous avons utilisé dans le modèle les poids de sondage de l'enquête, à savoir une pondération unique par observation. Ces poids sont employés dans une procédure d'estimation du pseudo MV, où les erreurs-types sont ajustées pour tenir compte de la pondération en utilisant un estimateur par linéarisation (Skinner, Holt et Smith 1989). Comme il s'agit de pondérations trimestrielles, elles ne conviennent pas pour estimer les totaux de population au niveau mensuel. Cependant, comme nous utilisons l'information provenant du registre pour tous les participants à l'EPA qui sont entrés dans l'enquête durant un trimestre particulier, ces pondérations sont appropriées pour l'estimation des modèles de Markov cachés.

## 4 Résultat pour les données de l'EPA et du PA appariées

En tout, nous avons estimé neuf modèles qui sont présentés au tableau 4.1. Tous ces modèles sont des modèles de Markov cachés d'ordre 1 avec deux variables indicatrices pour le type de contrat comme il est décrit à la section 3. Les probabilités d'erreur sont homogènes dans le temps. Les probabilités de transition (latentes) sont supposées être hétérogènes dans le temps; autrement dit, les logit des transitions peuvent dépendre du temps et du carré du temps. Ces modèles sont aussi des modèles de mélange finis qui comprennent trois classes latentes pour tenir compte de l'effet de l'hétérogénéité non observée dans l'état latent initial et dans les probabilités de transition latentes. Ce nombre de classes latentes a été choisi en comparant des variantes des modèles B'' et C comprenant différents nombres de classes latentes (les résultats de ces tests sont disponibles sur demande).

Les modèles A', A'' et A spécifient des erreurs de classification indépendantes (ECI) pour l'enquête, le registre et les deux ensembles de données, respectivement. Le modèle B' spécifie l'erreur dans l'enquête de façon qu'elle dépende des covariables  $V_{it}$  d'âge et d'interview par personne interposée, le modèle B'' spécifie des erreurs autocorrélées dans le registre, tandis que le modèle B combine ces deux spécifications. Les modèles C' et C'' étendent le modèle B'' en introduisant les prédicteurs  $Z_{it}$  (sexe, âge, niveau de scolarité et pays d'origine) pour les transitions, d'une part, et pour l'état initial ainsi que les transitions, d'autre part. Le modèle C étend le modèle B en introduisant les mêmes prédicteurs.

Le tableau 4.1 donne les valeurs de la log-vraisemblance, du critère d'information bayésien (BIC) et du critère d'information d'Akaike (AIC), ainsi que le nombre de paramètres pour neuf des modèles qui ont été estimés en utilisant les données de l'EPA et du PA appariées. Dans tous les modèles, les probabilités de transition (latentes) sont supposées être hétérogènes dans le temps; autrement dit, les logit des transitions peuvent dépendre du temps et du carré du temps.

Le modèle A spécifie que les données de l'enquête ainsi que les données du registre contiennent des erreurs de classification (indépendantes). Comme la qualité de l'ajustement de ce modèle aux données est meilleure que celle des modèles A' et A'' restreints, qui supposent que seule l'enquête (Modèle A') ou seul le registre (Modèle A'') contient des erreurs, nous concluons qu'il s'agit d'une preuve que les sources contiennent toutes deux des erreurs de classification.

Dans les modèles B', B'' et B, l'hypothèse ECI est relâchée pour l'enquête, pour le registre, ainsi que pour l'enquête et le registre, respectivement. Plus précisément, l'erreur de mesure dans les données d'enquête peut dépendre de l'âge de l'enquêté et du fait que l'information a été obtenue ou non par personne interposée, et l'erreur de mesure dans les données du registre peut dépendre du type de contrat latent et observé décalé. Cette dernière condition est obtenue en estimant un jeu distinct de probabilités d'erreur pour la répétition de *la même erreur* sur les diverses éditions de l'enquête ou du registre. Les versions restreintes du modèle B sont estimées également pour examiner si la violation de l'hypothèse ECI s'applique à l'erreur de mesure des données d'enquête seulement (modèle B') ou des données du registre seulement (modèle B''). Le fait que le modèle B'' soit mieux ajusté que les modèles A et B' indique que l'hypothèse ECI doit être relâchée pour l'indicateur de données du registre. Le modèle B améliore légèrement l'ajustement comparativement au modèle B'', ce qui indique que l'hypothèse ECI pour la variable indicatrice d'enquête doit aussi être relâchée dans un modèle sans prédicteurs pour les transitions et pour l'état initial.

**Tableau 4.1**  
**Mesures de l'ajustement pour huit modèles estimés au moyen de données de l'EPA et du PA appariées**

Modèle	Log-vraisemblance	BIC (LV)	AIC (LV)	Paramètres	$L^2$	dl	Valeur p
A': Enquête ECI	-286 814	574 118	573 716	44	240 543,4	69 327	1,6e-18 454
A'': Registre ECI	-454 196	908 882	908 480	44	575 307,7	69 327	8,5e-78 021
A: Les deux ECI	-284 413	569 384	568 926	50	235 742,1	69 321	4,8e-17 717
B': A + enquête non-ECI	-283 573	567 748	567 254	54	426 966,7	69 317	6,6e-50 302
B'': A + registre non-ECI	-246 054	492 732	492 220	56	435 025,8	69 315	2,9e-51 771
B: A + les deux non-ECI	-246 000	492 669	492 120	60	477 741,8	69 311	7,6e-59 639
C': B'' + prédicteurs des transitions	-245 282	491 590	490 748	92	486 186,8	69 279	1,8e-61 222
C'': B'' + prédicteurs état initial et transitions	-241 990	485 140	484 189	104	479 603,4	69 267	4,9e-60 003
C: B + prédicteurs état initial et transitions	-242 006	485 217	484 229	108	479 635,2	69 263	1,2e-60 010

Nota : Les modèles A', A'' et A spécifient des erreurs de classification indépendantes (ECI) pour l'enquête, le registre et les deux ensembles de données, respectivement. Le modèle B' spécifie une erreur dans l'enquête qui dépend de l'âge et de l'interview par personne interposée, le modèle B'' spécifie des erreurs autocorrélées dans le registre, tandis que le modèle B combine ces deux spécifications. Les modèles C' et C'' étendent le modèle B'' en introduisant le sexe, l'âge, le niveau de scolarité et le pays d'origine comme prédicteurs des transitions, ainsi que de l'état initial et des transitions, respectivement. Le modèle C étend le modèle B en introduisant les mêmes prédicteurs. Tous les modèles sont des modèles de mélange finis comportant trois classes latentes pour corriger l'hétérogénéité non observée dans l'état latent initial et dans les probabilités de transition latentes. En outre, tous les modèles supposent que les probabilités de transition latentes sont hétérogènes dans le temps. En particulier, nous conditionnons les probabilités de transition latentes sur une tendance linéaire pour le mois de l'observation ainsi que sur son carré.

Enfin, nous étendons les modèles B'' et B en incluant des covariables (sexe, âge, niveau de scolarité et pays d'origine) dans les modèles pour les probabilités de transition latentes et d'état initial latent (modèle C'' et C, respectivement). Le modèle C' est une version restreinte du modèle C''. Les variables explicatives ne peuvent avoir une incidence que sur les probabilités de transition latentes. Le fait que la qualité de l'ajustement du modèle C'' soit meilleure que celles du modèle B'' et du modèle C' indique que les covariables ont un effet significatif à la fois sur les transitions et sur les états initiaux. Le fait, d'après deux des trois mesures, que la qualité de l'ajustement du modèle C est moins bonne que celle du modèle C''

signifie que l'hypothèse d'ECI dans les données d'enquête devrait être gardée dans le modèle comprenant des variables explicatives pour les transitions et pour l'état initial (comme le montrent les résultats du modèle C, la taille de l'erreur de mesure dans les données d'enquête ne varie que légèrement avec l'âge et l'interview par personne interposée. Il s'agit d'une preuve supplémentaire en faveur du maintien de l'hypothèse d'ECI pour l'indicateur d'enquête. En fait, les estimations de la taille de l'erreur de mesure dans les données d'enquête ainsi que dans les données de registre et les estimations des probabilités de transition latentes sont très semblables pour les modèles C, C' et C''. Cela montre que les résultats de notre modèle sont robustes aux petites erreurs de spécification du modèle). Dans la suite de l'exposé, nous présentons les estimations calculées d'après le modèle C'' (les estimations calculées d'après les modèles C et C' sont disponibles sur demande).

Nous avons examiné diverses options de modèle non-ECI. En particulier, nous avons cherché à déterminer si l'erreur de mesure dans les données d'enquête diffère pour des secteurs où la mobilité de contrat et d'emploi est grande, comme le secteur de la santé, mais cela ne s'est pas avéré être le cas. Pour les données de registre, nous avons examiné des spécifications restreintes de rechange pour les erreurs corrélées, mais celles-ci ont donné des modèles dont la qualité de l'ajustement était pire que celle des modèles du tableau 4.1.

Examinons maintenant la quantité d'erreur de classification dans les deux sources de données. Selon l'équation 3.2, pour les données d'enquête et de registre, elle est représentée par les probabilités  $P(E_{it} = e_{it} | X_{it} = x_t)$  et  $P(C_{it} = c_{it} | X_{it} = x_t, X_{i(t-1)} = x_{t-1}, C_{i(t-1)} = c_{t-1})$ , respectivement. Les estimations d'après le modèle C'' sont présentées aux tableaux 4.2 et 4.3. En particulier, le tableau 4.2 montre que l'état de contrat permanent et l'état « autre » sont mesurés de façon très précise dans l'EPA, car presque toutes les personnes dans ces états sont classées correctement. Ce fait est indiqué par les grandes probabilités sur la diagonale principale du tableau. Une certaine erreur est observée pour les personnes qui possèdent en réalité un contrat d'emploi temporaire. 12,5 % de ces personnes déclarent qu'elles ont un contrat d'emploi temporaire, tandis que 4,2 % déclarent être dans une autre situation.

**Tableau 4.2**  
**Taille de l'erreur de mesure dans les données d'enquête selon le modèle C''**

Contrat latent en $t$	Contrat observé à la période $t$		
	Permanent	Temporaire	Autre
Permanent	0,998	0,001	0,002
Temporaire	0,125	0,832	0,042
Autre	0,004	0,005	0,991

Nota : Les erreurs-types sont toujours inférieures à 0,0001.

Le tableau 4.3 donne les probabilités estimées d'erreur de mesure pour les données de registre, qui, selon l'équation 3.2, dépendent des états observé et latent décalés. En raison des restrictions imposées (voir la section 3), nous avons estimé des paramètres (logit) d'erreur distincts pour la répétition de la même erreur entre les mois  $t - 1$  et  $t$ . Ces situations correspondent aux cellules ombrées dans le tableau 4.3. Comme le montre le tableau, les erreurs de mesure sont fortement autocorrélées; autrement dit, si une erreur a été commise au mois  $t - 1$  et qu'il était possible de répéter la même erreur (si la personne était restée dans le même état latent), l'erreur persistait presque certainement au mois  $t$ . Par exemple, si une personne possédant un contrat d'emploi permanent au mois  $t - 1$  était enregistrée par erreur comme ayant un contrat d'emploi temporaire et qu'elle possédait encore un contrat d'emploi



permanent au mois  $t$ , elle avait une probabilité de 0,968 d'être de nouveau enregistrée incorrectement comme ayant un contrat d'emploi temporaire au temps  $t$ . Pour les cinq autres erreurs possible, la probabilité d'une erreur de mesure persistante est un peu plus faible, mais n'est jamais inférieure à 0,84.

Un tableau différent se dégage lorsqu'aucune erreur n'est commise au temps  $t - 1$  ou quand une personne change d'état latent entre  $t - 1$  et  $t$ , et que, par conséquent, aucune répétition de l'erreur n'est possible. Dans ces cas, les données de registre sont presque exemptes d'erreur. Par exemple, quand une personne a été enregistrée correctement comme ayant un contrat d'emploi permanent au temps  $t - 1$  et qu'elle possède un contrat d'emploi temporaire au temps  $t$ , le type de contrat est enregistré correctement comme étant temporaire au temps  $t$  avec une probabilité de 0,930. En pratique, cela signifie que l'enregistrement initial du type de contrat est crucial pour le registre PA. Si cet enregistrement est correct, on peut avoir pleinement confiance dans le type de contrat enregistré pour la personne jusqu'à ce qu'un changement réel de situation sur le marché du travail ait eu lieu. En revanche, si le type de contrat de la personne est enregistré incorrectement au départ, cette erreur persistera presque certainement jusqu'à ce que le type de contrat de la personne change.

**Tableau 4.3**

**Probabilités conditionnelles d'erreur de mesure dans les données du registre selon le modèle C''**

Contrat observé en $t - 1$	Contrat latent en $t$	Contrat latent en $t - 1$	Contrat observé en $t$		
			Permanent	Temporaire	Autre
Permanent	Permanent	Permanent	0,986	0,009	0,004
Permanent	Permanent	Temporaire	0,986	0,009	0,004
Permanent	Permanent	Autre	0,986	0,009	0,004
Permanent	Temporaire	Permanent	0,045	0,930	0,025
Permanent	Temporaire	Temporaire	0,968	0,032	0,001
Permanent	Temporaire	Autre	0,045	0,930	0,025
Permanent	Autre	Permanent	0,005	0,005	0,990
Permanent	Autre	Temporaire	0,005	0,005	0,990
Permanent	Autre	Autre	0,913	0,000	0,087
Temporaire	Permanent	Permanent	0,027	0,973	0,000
Temporaire	Permanent	Temporaire	0,986	0,009	0,004
Temporaire	Permanent	Autre	0,986	0,009	0,004
Temporaire	Temporaire	Permanent	0,045	0,930	0,025
Temporaire	Temporaire	Temporaire	0,045	0,930	0,025
Temporaire	Temporaire	Autre	0,045	0,930	0,025
Temporaire	Autre	Permanent	0,005	0,005	0,990
Temporaire	Autre	Temporaire	0,005	0,005	0,990
Temporaire	Autre	Autre	0,001	0,842	0,157
Autre	Permanent	Permanent	0,039	0,000	0,961
Autre	Permanent	Temporaire	0,986	0,009	0,004
Autre	Permanent	Autre	0,986	0,009	0,004
Autre	Temporaire	Permanent	0,045	0,930	0,025
Autre	Temporaire	Temporaire	0,005	0,099	0,896
Autre	Temporaire	Autre	0,045	0,930	0,025
Autre	Autre	Permanent	0,005	0,005	0,990
Autre	Autre	Temporaire	0,005	0,005	0,990
Autre	Autre	Autre	0,005	0,005	0,990

Nota : Les erreurs-types sont systématiquement inférieures à 0,0001.

Afin d'estimer la quantité globale d'erreurs dans les données de registre, nous utilisons la probabilité a posteriori de posséder un type particulier de contrat latent à chaque point dans le temps. Cette probabilité

est estimée pour toutes les personnes comprises dans notre échantillon au moyen du modèle de Markov caché. Ces estimations sont assez exactes, car l'erreur de classification n'est que de 0,016. Les moyennes de ces probabilités sur l'ensemble des personnes et des points dans le temps sont présentées au tableau 4.4. En comparant les probabilités sur les diagonales principales des tableaux 4.1 et 4.4, nous voyons que l'erreur est plus grande dans l'indicateur de registre que dans l'indicateur d'enquête. En particulier, les personnes qui travaillent réellement aux termes d'un contrat temporaire ont une probabilité de 0,237 d'être enregistrées comme ayant un contrat permanent (0,125 dans les données d'enquête) et une probabilité de 0,079 d'être enregistrées comme se trouvant dans l'état « autre » dans le PA (0,042 dans les données d'enquête). Il existe également une erreur de classification pour les personnes qui possèdent réellement un contrat d'emploi permanent, car elles ont une probabilité de 0,081 d'être enregistrées comme des travailleurs temporaires et une probabilité de 0,031 d'être enregistrées comme étant dans une autre situation.

**Tableau 4.4**  
**Taille de l'erreur de mesure dans les données de registre selon le modèle C''**

Contrat latent en <i>t</i>	Contrat observé en <i>t</i>		
	Permanent	Temporaire	Autre
Permanent	0,888	0,081	0,031
Temporaire	0,237	0,684	0,079
Autre	0,032	0,017	0,951

Nota : Ces probabilités sont égales à la moyenne des probabilités a posteriori d'avoir un type particulier de contrat latent telles qu'elles sont estimées au moyen du modèle C'' avec une erreur de classification de 0,016.

Nous nous intéressons non seulement à l'erreur de mesure proprement dite, mais aussi à la grandeur de son effet sur l'estimation de la taille de l'emploi temporaire. En utilisant de nouveau la moyenne des probabilités a posteriori d'avoir un type particulier de contrat latent, nous estimons la taille de l'emploi temporaire aux Pays-Bas. Au tableau 4.5, nous comparons la taille de l'emploi temporaire estimée au moyen du modèle de Markov caché avec les distributions observées du type de contrat selon l'EPA et selon le PA. La probabilité a posteriori moyenne d'avoir un contrat temporaire est de 10,9 % et est comprise entre les valeurs obtenues d'après l'EPA et d'après le PA.

**Tableau 4.5**  
**Taille moyenne de l'emploi temporaire selon le modèle C''**

	Observée		Latente
	Enquête	Registre	
Permanent	0,667	0,597	0,634
Temporaire	0,087	0,130	0,109
Autre	0,246	0,273	0,257
Cas	48 297	174 480	174 480

Nota : Les probabilités latentes sont égales à la moyenne des probabilités a posteriori d'avoir un type particulier de contrat latent telles qu'elles sont estimées au moyen du modèle C'' avec une erreur de classification de 0,016.

Le tableau 4.6 donne l'évolution de la taille de l'emploi temporaire d'après les deux sources de données et d'après le modèle de Markov caché. Ce tableau confirme la constatation selon laquelle la taille de l'emploi temporaire calculée d'après notre modèle est comprise entre celles obtenues d'après les données du registre et d'après les données de l'enquête. On voit aussi que, durant la période de référence,

la proportion d'employés temporaires a augmenté. La faible baisse observée dans les données de registre en janvier 2008 (mois 13) comparativement à décembre 2007 (mois 12) s'explique par le fait que de nombreux contrats temporaires se terminent le 31 décembre et que, de surcroît, certains de ces contrats sont convertis en contrats d'emploi permanent. La fluctuation un peu plus importante de la taille de l'emploi temporaire selon les données d'enquête est due au fait que les participants à l'EPA sont interviewés tous les trois mois et que les diverses estimations mensuelles proviennent donc partiellement de différents participants à l'enquête.

Il est important d'examiner non seulement la variation agrégée, mais aussi la variation au niveau individuel, c'est-à-dire la probabilité d'une transition d'un emploi temporaire à un emploi permanent et inversement. Ces probabilités de transition sont présentées au tableau 4.7. En particulier, ce tableau donne les probabilités de transition latentes (moyennes) obtenues d'après le modèle C". Les probabilités de transition ont trait à une période de trois mois et une moyenne est calculée sur les 12 périodes de trois mois dans nos données. Si nous comparons les résultats du tableau 4.7 à ceux du tableau 2.4, nous voyons que les probabilités de transition latentes sont beaucoup plus faibles que celles découlant des données de registre et des données d'enquête. D'après les probabilités de transition latentes, 3,2 % des personnes ayant un contrat d'emploi temporaire possédaient un contrat d'emploi permanent trois mois plus tard, mais selon les données d'enquête et de registre, ces pourcentages étaient de 5,7 % et 8,5 %, respectivement. Cela montre que l'erreur de mesure donne lieu à une inflation de la taille des probabilités de transition. Une telle inflation serait clairement attendue quand les erreurs sont indépendantes dans le temps (Hagenaars 1990, 1994). Si les erreurs ne sont pas indépendantes dans le temps, comme dans notre cas, l'attente est moins claire, car les erreurs pourraient soit accroître soit réduire les probabilités de transition, selon la nature et la taille de l'association. La même tendance à la sous-estimation de la stabilité peut être observée pour l'état de contrat permanent : 98,1 % et 96,7 % de personnes sont restées dans cet état selon les données de l'enquête et les données du registre, respectivement, alors que la stabilité réelle était de 98,7 %.

**Tableau 4.6**

**Évolution de la proportion d'employés temporaires durant la période de janvier 2007 à mars 2008**

Mois	Source		
	Enquête	Registre	Latente
1	0,080	0,123	0,102
2	0,082	0,124	0,103
3	0,085	0,123	0,102
4	0,084	0,128	0,103
5	0,084	0,129	0,103
6	0,090	0,129	0,104
7	0,089	0,130	0,105
8	0,087	0,131	0,106
9	0,091	0,135	0,110
10	0,087	0,134	0,112
11	0,088	0,135	0,114
12	0,091	0,135	0,114
13	0,090	0,131	0,116
14	0,089	0,131	0,118
15	0,096	0,132	0,121

Nota : Les données d'enquête comprennent des observations trimensuelles par personne, tandis que les données de registre comprennent des observations mensuelles par personne. Les probabilités latentes sont égales à la moyenne des probabilités a posteriori d'avoir un type particulier de contrat latent telles qu'elles sont estimées au moyen du modèle C" avec une erreur de classification de 0,016.

**Tableau 4.7**  
**Transitions observées sur 3 mois dans l'EPA et le PA et transitions latentes selon le modèle C**

Transitions latentes				
		Permanent	Temporaire	Autre
Contrat en t-3	Permanent	0,987	0,006	0,007
	Temporaire	0,032	0,931	0,037
	Autre	0,009	0,030	0,961
	Total	0,634	0,110	0,256

Nota: Les probabilités latentes sont égales à la moyenne des probabilités a posteriori d'avoir un type particulier de contrat latent telles qu'elles sont estimées au moyen du modèle C'' avec une erreur de classification de 0,016.

## 5 Conclusion

Dans le présent article, nous avons étudié l'erreur de mesure du type de contrat d'emploi dans l'EPA des Pays-Bas en appariant les données de sa composante longitudinale pour la période de 2007 au début de 2008 à un ensemble de données de registre unique, c'est-à-dire le PA. Nous avons appliqué plusieurs modèles de Markov cachés dans lesquels le type de contrat réel est traité comme un état latent et dans lesquels les données provenant de l'enquête et du registre servent d'indicateurs observés du contrat réel d'une personne. Nous avons modélisé l'erreur de mesure dans les deux sources de données en tenant compte du fait que l'erreur dans le registre est corrélée entre les périodes.

Nos résultats montrent que les données du registre contiennent plus d'erreurs que les données d'enquête et qu'elles ne peuvent donc pas être utilisées comme normes de référence. Cependant, l'amélioration de l'enregistrement initial dans les données du registre peut accroître considérablement leur qualité, car l'erreur de mesure dans l'indicateur du type de contrat provenant de cet ensemble de données est autocorrélée.

L'erreur de mesure résulte en une sous-estimation du pourcentage de personnes possédant un contrat d'emploi temporaire. Dans l'EPA, ce pourcentage est de 8,9 %, mais après correction de l'erreur de mesure, il passe à 10,9 %. Un autre effet de l'erreur de mesure est qu'elle produit des probabilités de transition fortement surestimées. Selon l'EPA et le PA, la probabilité de transition entre l'emploi temporaire et l'emploi permanent pour une période de trois mois est de 5,7 % et de 8,5 %, respectivement, tandis que la probabilité de transition latente correspondante n'est que de 3,2 %. Cette constatation est particulièrement importante pour les responsables de l'élaboration des politiques aux Pays-Bas, car elle indique clairement que la mobilité de l'emploi temporaire vers l'emploi permanent est nettement moindre qu'on ne le pensait au départ.

Les résultats de la présente étude demeurent relativement stables pour les diverses spécifications de modèle que nous avons évaluées. Cela montre qu'ils sont robustes aux petites erreurs de spécification du modèle. Cependant, ils restent quelque peu dépendants des hypothèses du modèle. D'autres applications et tests de sensibilité permettront de vérifier plus en profondeur la validité de nos résultats. De futures études pourraient se concentrer sur des tests de sensibilité en faisant appel à des simulations Monte Carlo.

## Remerciements

Les auteurs sont reconnaissants à l'endroit de *Statistics Netherlands* pour l'accès aux données de cet article. Les auteurs remercient aussi Frank van der Pol, Wendy Smits, Ruben van Gaalen et les participants aux conférences de l'ESPE et de l'EALE ainsi que les participants du groupe de recherche SILC de l'Université libre d'Amsterdam pour leurs pertinents commentaires et suggestions. La contribution de Jeroen Vermunt a été supportée par la *Netherlands Organization for Scientific Research* (NWO) [Numéro de subvention VICI 453-10-002].

## Bibliographie

- Abowd, J.M., et Stinson, M.H. (2005). Estimating measurement error in SIPP annual job earnings: A comparison of census survey and SSA administrative data. Papier technique, *U.S. Census Bureau*.
- Bartolucci, F., Lupporelli, M. et Montanari, G.E. (2009). Latent markov model for longitudinal binary data: An application to the performance evaluation of nursing homes. *Annals of Applied Statistics*, 3(2), 611-636.
- Bassi, F., Hagenaars, J.A., Croon, M.A. et Vermunt, J.K. (2000). Estimating true changes when categorical panel data are affected by uncorrelated and correlated classification errors. *Sociological Methods and Research*, 29(2), 230-268.
- Baum, L.E., Petrie, T., Soules, G. et Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. *The Annals of Mathematical Statistics*, 41(1), 164-171.
- Bentolila, S., et Bertola, G. (1990). Firing costs and labour demand: How bad is eurosclerosis? *The Review of Economic Studies*, 57(3), 381-402.
- Biemer, P. (2004). Une analyse de l'erreur de classification pour les questions sur l'emploi révisées de la Current Population Survey. *Techniques d'enquête*, 30, 2, 141-155.
- Biemer, P. (2011). *Latent Class Analysis of Survey Error*. New Jersey : John Wiley & Sons, Inc.
- Biemer, P.P., et Bushery, J.M. (2000). Validité de l'analyse markovienne de structure latente pour l'estimation de l'erreur de classification des données sur la population active. *Techniques d'enquête*, 26, 2, 157-171.
- Bollinger, C.R. (1996). Bounding mean regressions when a binary regressor is mismeasured. *Journal of Econometrics*, 73(2), 387-399.
- Booth, A.L. (1997). An analysis of firing costs and their implications for unemployment policy. Dans *Unemployment Policy*, (Éds., D.J. Snower et G. de la Dehesa). Cambridge : Cambridge University Press.
- Bound, J., Brown, C., Duncan, G.J. et Rodgers, W.L. (1994). Evidence on the validity of cross-sectional and longitudinal labor market data. *Journal of Labor Economics*, 12(3), 345-368.

- Bound, J., Brown, C. et Mathiowetz, N. (2001). Measurement error in survey data. Dans *Handbook of econometrics*, (Éds., J.J. Heckman et E. Leamer), Amsterdam : Elsevier, 5, 3705-3843.
- Brown, C., et Medoff, J.L. (1996). Employer characteristics and work environment. *Annales D'Économie et de Statistique*, 41, 275-298.
- Cahuc, P., et Postel-Vinay, F. (2002). Temporary jobs, employment protection and labor market performance. *Labour Economics*, 9(1), 63-91.
- Dempster, A.P., Laird, N.M. et Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39(1), 1-38.
- Duncan, G.J., et Hill, D.H. (1985). An investigation of the extent and consequences of measurement error in labor-economic survey data. *Journal of Labor Economics*, 3(3), 508-522.
- Gottschalk, P. (2005). Downward nominal-wage flexibility: Real or measurement error. *Review of Economics and Statistics*, 87(3), 556-568.
- Gottschalk, P., et Huynh, M. (2010). Are earlings inequality and mobility overstated? The impact of non-classical measurement error. *Review of Economics and Statistics*, 92(2), 302-315.
- Hagenaars, J.A. (1990). *Categorical Longitudinal Data Log-Linear Panel, Trend and Cohort Analysis*. Newbury Park, CA : Sage Publications.
- Hagenaars, J.A. (1994). Latent variables in log-linear models of repeated observations. Dans *Latent Variable Analysis: Applications for Developmental Research*, (Éds., A. von Eye et C.C. Clogg). Thousand Oaks, CA : Sage Publications, 329-352.
- Heckman, J.J., et Singer, B.L. (1984). A method for minimising the impact of distributional assumptions in econometric models for duration data. *Econometrica*, 52 (2), 271-320.
- Hilbers, P., Houwing, H. et Kösters, L. (2011). De flexibele schil – overeenkomsten en verschillen tussen CBS- en UWV-cijfers [the flexible periphery – similarities and differences between CBS and UWV-data]. Dans *Socialeconomische Trends, 2<sup>e</sup> Kwartaal 2011 [Socioeconomic Trends, 2<sup>nd</sup> Trimester 2011]*, (Éds., B. Hermans et coll.). Den Haag/Heerlen : Statistics Netherlands, 26-33.
- Kapteyn, A., et Ypma, J.Y. (2007). Measurement error and misclassification: A comparison of survey and register data. *Journal of Labor Economics*, 25(3), 513-551.
- Langeheine, R. (1994). Latent variable markov models. Dans *Latent Variables Analysis. Applications for Developmental Research*, (Éds., A. von Eye et C. Clogg). Thousand Oaks, Californie : Sage Publications, 373-395.
- Manzoni, A., Vermunt, J.K., Luijkx, R. et Muffels, R. (2010). Memory bias in retrospectively collected employment careers: A model-based approach to correct for measurement error. *Sociological Methodology*, 40(1), 39-73.
- Mars, G. (2011, décembre). *Cijfers over Flexibele Arbeidsrelaties - Confrontatie Van Bronnen en Defities [Figures on Flexible Labour Relations - Confrontation of Sources and Definitions]*. Statistics Netherlands, rapport nr SAH-2011-H11. La Haye/Heerlen.

Mathiowetz, N.A. (1992). Errors in reports of occupations. *Public Opinion Quarterly*, 56(3), 352-355.

OECD (2002). *Employment Outlook 2002*. Paris : Auteur.

OECD (2012). *Country Statistical Profiles*. Base de données de l'OECD : consultée le 16/12/2012 à partir de <http://stats.oecd.org/>.

Paas, L.J., Vermunt, J.K. et Bijmolt, T.H. (2007). Discrete-time discrete-state latent markov modelling for assessing and predicting household acquisitions of financial products. *Journal of the Royal Statistical Society, Series A*, 170(4), 955-974.

Pavlopoulos, D., Muffels, R. et Vermunt, J.K. (2012). How real is mobility between low pay, high pay and non-employment. *Journal of Royal Statistical Society, Series A*, 175(3), 749-773.

Pischke, J.-S. (1995). Measurement error and earnings dynamics: Some estimates from the PSID validation study. *Journal of Business and Economic Statistics*, 13(3), 305-314.

Poulsen, C.S. (1990). Mixed markov and latent markov modelling applied to brand choice behaviour. *International Journal of Research in Marketing*, 7(1), 5-19.

Rendtel, U., Langeheine, R. et Berntsen, R. (1998). The estimation of poverty dynamics using different measurements of household income. *Review of Income and Wealth*, 44(1), 81-98.

Rodgers, W.L., Brown, C. et Duncan, G.J. (1993). Errors in survey reports of earnings, hours worked, and hourly wages. *Journal of the American Statistical Association*, 88(3), 345-368.

Sels, L., et Van Hootegeem, G. (2001). Seeking the balance between flexibility and security: A rising issue in the low countries. *Work, Employment and Society*, 15(2), 327-352.

Shorrocks, A.F. (1976). Income mobility and the markov assumption. *Economic Journal*, 86, 566-578.

Skinner, C.J., Holt, D. et Smith, T.M.F. (1989). *Analysis of Complex Surveys*. Wiley.

van der Pol, F., et Langeheine, R. (1990). Mixed markov latent class models. *Sociological Methodology*, 20, 213-247.

Vermunt, J.K. (1997). *Log-Linear Models for Event Histories*. Londres : SAGE publications.

Vermunt, J.K., et Magidson, J. (2008). *LG - Syntax User's Guide: Manual for Latent GOLD 4.5 Syntax Module*. Belmont Massachusetts : Statistical Innovations Inc.

Vermunt, J.K., Langeheine, R. et Böckenholt, U. (1999). Discrete-time discrete-state latent markov models with time-constant and time-varying covariates. *Journal of Educational and Behavioral Statistics*, 24, 178-205.

Vermunt, J.K., Tran, B. et Magidson, J. (2008). Latent class models in longitudinal research. Dans *Handbook of Longitudinal Research: Design, Measurement, and Analysis*, (Éd., S. Menard). Burlington, MA : Elsevier, 373-385.